
Several nodulins of soybean share structural domains but differ in their subcellular locations

Fred A.Jacobs¹, Mingda Zhang, Marc G.Fortin and Desh Pal S.Verma*

Centre for Plant Molecular Biology, Department of Biology, McGill University, Montreal, H3A 1B1, Canada

Received August 21, 1986; Revised and Accepted December 11, 1986

ABSTRACT

Four soybean cDNA nodule-specific clones encoding nodulin-23, -26b, -27 and -44 were observed to cross-hybridize under low stringency conditions. Nucleotide sequence analysis revealed that the cDNAs contain three distinct domains: two domains with 70 to 95% homology separated by a third domain unique to each cDNA. Despite a number of nucleotide insertions and deletions, the protein sequences are conserved in the two domains which correlate with the homologous nucleotide domains. The amino terminal domain of each nodulin contains putative signal sequences for membrane translocation, although only two (nodulin-23 and -44) meet all the criteria for a functional signal. Immuno-precipitation of hybrid-release translation products of the four cDNAs revealed that nodulin-23 is associated with the peribacteroid membrane while nodulin-27 is in the cytoplasmic fraction of the nodule. These four nodulins are members of a diverse family with conserved structural features and the genes encoding them appear to have recently evolved from a common ancestor.

INTRODUCTION

Successful infection of leguminous plants by *Rhizobium* spp. leads to the development of a specialized plant organ, the root nodule (for reviews, see 1,2). The symbiotic relationship which ensues allows the reduction of dinitrogen and its assimilation by the host while providing an ecological niche for the microsymbiont. A number of genes are specifically expressed in the nodule, encoded both by host and microsymbiont genomes. The products of these genes, called nodulins, have been characterized from a number of legumes (3-5), however, the functions of only a few have been determined: leghemoglobin (6), uricase II (nodulin-35) (7) and sucrose synthetase (nodulin-100) (8). Nodulin gene expression is not only tissue-specific but also cell type-specific (9). A number of nodulins have been identified to be involved in the formation of the subcellular compartment enclosing the microsymbiont (10). Little is known of the evolution and regulation of the apparently large number of nodule-specific genes (11,12).

Preliminary observations have indicated that several nodule-specific cDNA

clones hybrid-select a number of related mRNA species encoding different molecular weight nodulins and thus appear to be members of an abundantly transcribed multigene family (12,13,19). The nucleotide sequences of several of these clones were analyzed and were found to have common structural features. Their sequence homologies extend into the 3' non-coding region, which suggests that these sequences may have recently evolved from a common ancestor. Despite the similarity in their structures, these nodulins are located in different subcellular compartments.

MATERIALS AND METHODS

Nodulin cDNA Clones

A full length cDNA clone for nodulin-26b was obtained from a nodule specific cDNA library of Glycine max (cv. Prize) as described previously (12). Clone 45 (12) was used to isolate a full length cDNA clone for nodulin-27. Clone 15 (encoding nodulin-44) and clone 25 (encoding nodulin-23) have been previously characterized (12,14).

DNA Sequencing

Full length cDNA clones for nodulin-26b and -27 were sequenced following digestion with various restriction enzymes and cloning into the M13 vectors mp18 and mp19. The dideoxynucleotide chain termination method was used for sequencing the DNA with [^{32}S]dATP- αS and [$\alpha^{32}\text{P}$]dCTP (14). The primary structures for nodulin-23 and nodulin-44(E27) have been previously described (14,16).

Computer Analysis

Nucleotide sequence assembly and analysis were performed with the Pustell Sequence Analysis programs from International Biotechnologies, Inc. Computer analysis of cDNA-derived protein sequences were performed with the FastP programs (17).

Hybrid-release Translation

Total poly(A)⁺ RNA from soybean nodules were prepared for hybrid-release translation (HRT) (18). Briefly, liquid nitrogen frozen nodules were ground and homogenized in buffered 4M guanidinium isothiocyanate before the nucleic acids were precipitated with isopropyl alcohol. The precipitate was solubilized in 8M urea and the RNA was precipitated with LiCl followed by oligo(dT)-cellulose chromatography.

Specific nodulin mRNAs were translated in vitro after hybrid-selection (12). Electrophoretically purified cDNAs for nodulin-23, -26b, -27 and -44 were denatured and bound to 0.2 cm² pieces of Gene Screen (New England

Nuclear). Each filter was hybridized with 10 ug poly(A)⁺ RNA, washed and the mRNA was eluted with water at 95°C. The recovered mRNA was then translated in a rabbit reticulocyte-lysate (Bethesda Research Laboratories, MD) with [³⁵S] methionine.

Immuno-precipitation and Electrophoresis of Translation Products

HRT products were immuno-precipitated with antisera raised against soluble (S100) nodule specific (3) or peribacteroid membrane specific material and subjected to SDS-PAGE (10).

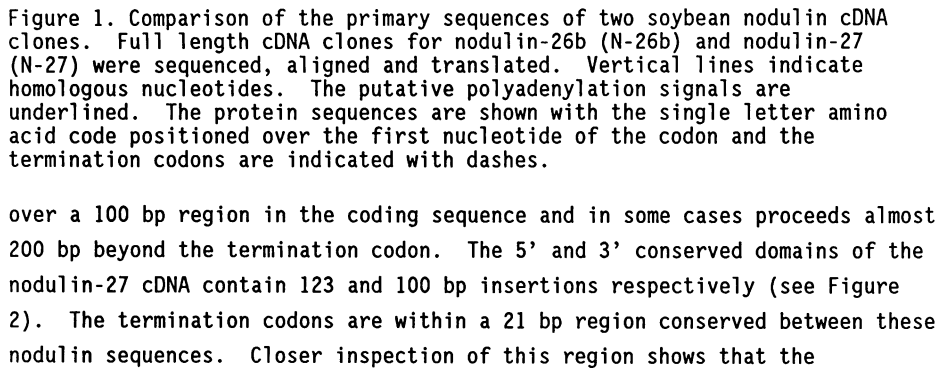
RESULTS

Primary Structures of Nodulin-26b and Nodulin-27

Complete nucleotide sequences of the nodulin-26b and -27 cDNAs were aligned and are displayed in Figure 1. Several domains of homology between these two sequences are observed, starting from the initiation codon. Although specific areas of mismatches occur, the overall homology in both the coding and non-coding homologous regions is greater than 80%. The two sequences are not co-linear as several unique deletions and/or insertions exist in both cDNAs. Of particular note are the two separate codon deletions (at positions 259 and 372 in nodulin-27), and two single bp insertions or deletions, at positions 586 and 543 in nodulin-26b and -27 respectively, which temporarily shift the reading frames out of phase for 25 bp.

The position of the initiation codons of nodulin-26b and -27 are different, although the sequences are homologous. The ATG used as the initiation codon in nodulin-27 exists in nodulin-26b, but in the latter case, the reading frame is almost immediately terminated. Instead, nodulin-26b is initiated by an unique ATG 16 bp downstream. The termination codons are also unique. The sequences immediately preceding the termination codon of nodulin-26b are non-homologous, with the homology resuming after the termination codon.

A tripartite nature of the nodulins' domain structure becomes the dominant feature when the comparisons of the nodulins is extended between nodulin-23, -26b and -27 with nodulin-44(E27) (Figure 2.) Partial sequencing of the nodulin-44 cDNA (clone 15) obtained from our cDNA library revealed that it was almost identical to the nodulin-E27 cDNA sequence of Sengupta-Gopalan *et al.* (16). The first homologous domain in these four nodulins is an almost uniform 250 bp in length, with the initiation codon found within a 26 bp region at the 5' border. The second homologous domain, however, is not so clearly defined. In particular, this domain commences



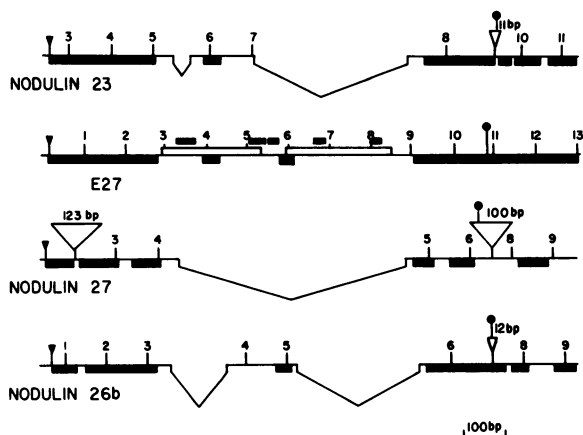


Figure 2. Comparison of nodulin cDNA sequences. The full length cDNA sequences of nodulin-23 (14), -26b and -27 (Figure 1) were compared with that of nodulin-44(E27) (designated E27) (16). Regions of >80% homology with nodulin-44(E27) are indicated with solid bars below the lines, with the translational start and stop codons shown with closed triangles and circles respectively. Open and closed bars above the nodulin-44(E27) cDNA sequence display internal recursive elements. Insertions are indicated by open triangles.

nodulin-26b, -27 and -44(E27) termination codons are different. Significant sequence homology exists around the termination codons, particularly between nodulin-27 and -44(E27), where the same nine nucleotide sequence codes for two termination codons in the different reading frames. The termination codons in these nodulins therefore appear to be generated by independent events. The termination codons of nodulin-23 and -26b are identical.

Structural Analysis of Nodulin Polypeptides

The conservation in domains of the nodulin cDNAs is reflected at the amino acid sequence level (Figure 3). Despite of a number of nucleotide insertions and/or deletions, the reading frames are maintained. These domains include up to 30% conservatively substituted amino acids. However, their secondary structures and hydrophobicity profiles of these polypeptides, as predicted by the algorithms described in (19,20), are markedly dissimilar. Nodulin-23 possesses an α helix domain around residue 78 which, according to its hydrophobic moment plot (20), is a potential trans-membrane structure.

Analysis of the amino terminal sequences of these nodulins reveal characteristics associated with signal peptides for membrane translocation (Figure 4) (see ref. 21). However, only nodulin-23 and -44(E27) meet all the

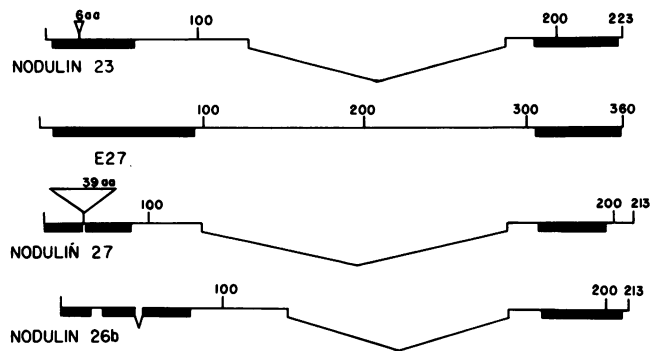


Figure 3. Comparison of nodulin amino acid sequences. The peptide sequences of nodulin-23, -26b, -27 and -44(E27) (designated E27) were determined from the longest open reading frames and compared. Regions of >80% homology, including conservatively substituted amino acids, are indicated by solid bars below the lines. Insertions are indicated by open triangles.

criteria of a functional signal sequence. Nodulin-26b lacks the first six amino acid residues of the signal sequence and nodulin-27 contains a lysine residue at the -5 position of a potential cleavage site, a residue not normally found at this location (21). No sequence homology between these putative signal peptides has been found with nodulin-24, which has an *in vitro* cleavable signal peptide (22). These interpretations lead us to predict that nodulin-23 and -44(E27) are membrane associated while nodulin-26b and -27 are in the soluble fraction of the cell.

Subcellular Location of Nodulins

In order to confirm the predicted subcellular locations of these nodulins, translation products of hybrid-selected nodulin mRNAs were immuno-precipitated with specific antibodies and subjected to SDS-PAGE

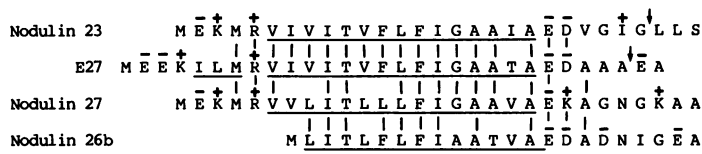


Figure 4. Putative amino terminal signal sequences in nodulins. The amino terminal peptide sequences of nodulin-23, -26b, -27 and -44(E27) were aligned. Vertical lines indicate direct amino acid homology of nodulin-23, -26b and -27 with nodulin-44(E27). Basic and acidic amino acids are shown with '+' and '-' respectively while the hydrophobic regions are underlined. The arrows indicate putative sites for signal peptide cleavage.

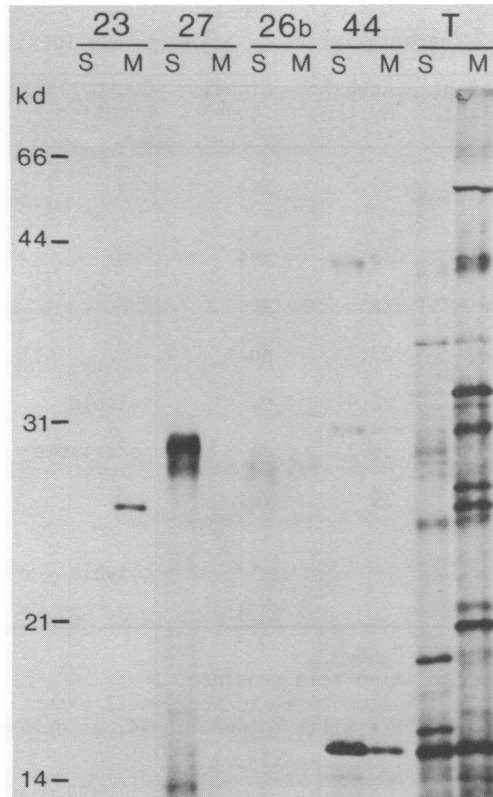


Figure 5. Electrophoresis of immuno-precipitated hybrid-release translation products of nodulin mRNAs. Specific hybrid selected mRNAs for nodulin-23, -26b, -27 and -44 were translated *in vitro* and precipitated with antisera raised against either soluble (S100) nodule specific (S) or peribacteroid membrane (M) material. 'T' indicates translation products from total nodule poly(A)⁺RNA. The immuno-precipitated material was resolved by SDS-PAGE and autofluorography .

(Figure 5). Nodulin-23 and -27 reacted specifically with antisera raised against peribacteroid membrane and soluble (S100) nodule specific material respectively. These findings suggest that nodulin-23 and -27 are located respectively in the peribacteroid membrane and in the cytosolic compartment of the nodule. Nodulin-26b and -44, however, reacted weakly with both antisera, thus their subcellular locations could not be precisely deduced. The absence of reactivity of nodulin-44 with the peribacteroid membrane antisera is not inconsistent with our predictions since the protein could be associated with other endo-membranes. The nodulin-44 cDNA clone appears to

Table 1.
Characteristics of Soybean Nodulins

Nodulin	Apparent ¹ MW (kD)	Actual ² MW (kD)	Signal Sequence	Sub-cellular Location	Function
23 (C51) ³	25	24.3	yes	PBM	-
24	24	15.1	yes	PBM	-
26	26.5	22.5	yes	PBM	-
26b	25.5	23.5	no	-	-
27	27	22.4	no	soluble	-
35	33	35.1	no	peroxisome	Uricase II
44 (E27) ⁴	42	39	yes	-	-
100	90	-	-	soluble	Sucrose Synthetase

1) determined by SDS-PAGE.

2) derived from the amino acid sequence.

3) nodulin-C51 (16) is identical to nodulin-23 except for 3 nucleotides.

4) nodulin-E27 (16) is similar to nodulin-44, as determined by partial sequence analysis.

hybrid-select at least two other mRNA species in addition to the one coding for the 42 kD peptide. However, these additional products become apparent only following immuno-precipitation (c.f. ref.10).

Table 1 summarizes the characteristics of these and other nodulins in addition to leghemoglobin analyzed to date. Several of the abundantly transcribed nodulins are integral parts of the peribacteroid membrane. This includes nodulin-26, which has no homology with the nodulin family described here (23). The apparent molecular weights of most of the nodulins is larger than the actual molecular weights, which appears to be due to specific secondary structural features (see ref. 22).

DISCUSSION

We have described the structures of four nodulins which appear to be encoded by an unique gene family. These nodulins have three domains, two homologous domains separated by a third non-homologous domain of variable length. Thirty percent of the amino acid sequence of the homologous domains,

however, consists of conservatively substituted amino acids. Also, the initiation and termination codons are approximately coincident while small insertions and deletions and transient reading frame shifts occur. The second homologous domain is notable for the significant extension of the cDNA homology into the 3' non-coding region. These findings suggest that the four nodulins described are closely related and have recently evolved from a common ancestor. The divergence of the peptide sequences indicate that these proteins may not have essential enzymatic functions. However, because of the overall conservation in their structures, including the conserved positions of the uniquely encoded initiation and termination codons, these nodulins must be subjected to specific constraints. Searches through the National Biomedical Protein Data Bank have not revealed significant homologies with any known protein sequence.

These four nodulins are only a subset of a larger family of related nodulins. This is evident from the multiple bands of nodulin-44 and the double bands of nodulin-26b and -27 cDNA HRT products immuno-precipitated and resolved by SDS-PAGE. Multiple bands in Southern and Northern blots probed with these cDNAs have also been detected (9).

The members of this family encode polypeptides of different molecular weights with different subcellular locations. This contrasts with most other gene families which code for products which are similar, such as those encoding leghemoglobins (24) and actins (25). This nodulin family may have both expressed and silent genes. We have isolated a number of different genomic clones which hybridize with the nodulin-26b cDNA under low stringency (unpublished data). The diversity of these sequences may have resulted from a series of gene duplications and acquisition of foreign sequences to provide a number of possible candidates for nodule-specific functions, none of which may have yet been optimally configured during the short evolutionary period of nitrogen-fixing legumes. The rapid divergence of this large family may then provide a pool of different nodulins, some of which may eventually become more efficient. The intensive breeding to which Glycine max cultivars were subjected in the past may also have increased the sequence diversity. The origin of this unique family may be from the host response of genomic stress to early pathogenic interacts with Rhizobium spp.

ACKNOWLEDGEMENTS

We are grateful to C. Sengupta-Gopalan for providing the cDNA sequences for nodulin-C51 and -E27 prior to publication. This work was supported by the National Science and Engineering Research Council of Canada (NSERC). MGF and MZ were supported with fellowships from NSERC and the CIDA McGill/Mankai Biotechnology project.

Abbreviations: bp: base pair, HRT: hybrid release translation, kD: kilo Dalton, SDS-PAGE: Sodium dodecyl sulphate polyacrylamide gel electrophoresis

¹Current address: Biotechnology Research Institute, Montreal, Canada

*To whom correspondence should be addressed

REFERENCES

1. Verma, D.P.S. and Long, S. (1983) in International Review of Cytology, K Jeon, Ed. (Academic Press, New York), Suppl. 14, pp 211-245.
2. Verma, D.P.S. and Nadler, S. (1984) in Genes Involved in Microbe-Plant Interactions. DPS Verma and T Hohn, Eds. (Wein New York: Springer-Verlag), pp 58-84.
3. Legocki, R.P. and Verma, D.P.S. (1980) Cell 20, 153-163.
4. Govers, S., Gloudemans, T., Moerman, M., van Kammen, A. and Bisseling, T. (1985) The EMBO J. 4, 861-867.
5. Lang-Unnash, N. and Ausubel, F.M. (1985) Plant Physiol. 77, 833-839.
6. Appleby, C.A., Nicola, N.A., Hurrell, J.G.R. and Leach, D.J. (1975) Biochem. 14, 4444-4450.
7. Nguyen, T., Zelechowska, M., Foster, V., Bergman, H. and Verma, D.P.S. (1985) Proc. Natl Acad. Sci. USA 82, 5040-5044.
8. Thummler, F. and Verma, D.P.S. (1986) in preparation.
9. Verma, D.P.S., Fortin, M.G., Stanley, J., Mauro, V.P., Purohit, S. and Morrison, N. (1986) Plant Mol. Biol. 7, 51-61.
10. Fortin, M.G., Zelechowska, M. and Verma, D.P.S. (1985) The EMBO J. 4, 3041-3046.
11. Auger, S. and Verma, D.P.S. (1981) Biochem. 20, 1300-1306.
12. Fuller, F., Kunstner, P.W., Nguyen, T. and Verma, D.P.S. (1983) Proc. Natl Acad. Sci. USA 80, 2594-2598.
13. Fuller, F. and Verma, D.P.S. (1984) Plant Mol. Biol. 3, 21-28.
14. Mauro, V.P., Nguyen, T., Katinakis, P. and Verma, D.P.S. (1985) Nucl. Acids Res. 13, 239-249.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl Acad. Sci. USA 74, 5463-5467.
16. Sengupta-Gopalan, C., Pitas, J.W., Thompson D.V. and Hoffman, L.M. (1986) Molec. Gen. Genet. 203, 410-420.
17. Lipman, D.J. and Pearson, W.R. (1985) Science 227, 1435-1441.
18. Clemens, M.J. (1984) in Transcription and Translation, A Practical Approach. BD Hames and SJ Higgins Eds, (Oxford, IRL Press), pp 211-270.
19. Argos, P., Narayana, S.V.L. and Nielson, N.C. (1985) The EMBO J. 4, 1111-1117.
20. Eisenberg, D. (1984) Ann. Rev. Bioch. 53, 595-623.
21. von Heijne, G. (1986) Nucl. Acids Res. 14, 4683-4690.
22. Katinakis, P. and Verma, D.P.S. (1985) Proc. Natl Acad. Sci. USA 82, 4157-4161.
23. Fortin, M.G., Morrison, N. and Verma D.P.S. (submitted).
24. Lee, J. and Verma D.P.S. (1984) The EMBO J. 3, 2745-2752.
25. Korn, E.D. (1982) Physiol.Rev. 62, 672-737.